Available online at www.sciencedirect.com

ScienceDirect

Vision Research 47 (2007) 219–230

Vision Research

www.elsevier.com/locate/visres

Spatial grouping in human vision: Temporal structure trumps temporal synchrony

Sharon E. Guttman *, Lee A. Gilroy, Randolph Blake

Department of Psychology, Vanderbilt University, 301 Wilson Hall, 111 21st Ave. South, Nashville, TN 37203, USA

Received 10 April 2006; received in revised form 31 August 2006

Abstract

Temporal information promotes visual grouping of local image features into global spatial form. However, experiments demonstrating time-based grouping typically confound two potential sources of information: *temporal synchrony* (precise timing of changes) and *temporal structure* (pattern of changes over time). Here, we show that observers prefer temporal structure for determining perceptual organization. That is, human vision groups elements that change according to the same global pattern, even if the changes themselves are not synchronous. This finding prompts an important, testable prediction concerning the neural mechanisms of binding: patterns of neural spiking over time may be more important than absolute spike synchrony.

Keywords: Visual grouping; Temporal structure; Spatial structure; Synchrony; Binding problem

1. Introduction

Looking around the visual world, we readily perceive distinct, meaningful objects. Remarkably, however, the initial stages of visual processing register only local image features comprising those objects. These local, spatially distributed features must be grouped into coherent, global objects that are segmented from one another and from the backgrounds against which they appear. Called the *binding problem* in contemporary parlance, the importance of this grouping operation for figure–ground organization first was highlighted by Gestalt psychologists in the early part of the previous century, and it remains a central problem in vision science today.

Among the sources of stimulus information indicating whether spatially distributed elements should be grouped, the role of temporal factors has been of enduring interest. Numerous psychophysical studies indicate that the visual system can capitalize on the constraint that, in the natural environment, visual features that change at the same times likely belong to a single object (e.g., Alais, Blake, & Lee, 1998; Guttman, Gilroy, & Blake, 2005; Kandil & Fahle, 2001; Lee & Blake, 1999, 2001; Sekuler & Bennett, 2001; Suzuki & Grabowecky, 2002; Usher & Donnelly, 1998). While there is disagreement concerning specifics of the neural operations that promote temporal grouping, this body of research converges on the notion that the visual system exploits temporally coincident change, broadly construed, as a cue for stimulus binding (see Blake & Lee, 2005, for a review of alternative hypotheses).

To date, research on temporal correlation as a binding agent has primarily focused on *temporal synchrony*. Multiple events occurring over time are synchronous—and thus theoretically will be bound—if the individual events occur at the same moments in time. Synchrony, however, is not the only conceivable temporal signature for grouping. One might also envision feature grouping based on *common temporal structure* among those features. Temporal

* Corresponding author. Present address: Department of Psychology, MTSU, P.O. Box 87, 1301 E. Main Street, Murfreesboro, TN 37132, USA. Fax: +1 615 898 5207.

E-mail address: sguttman@mtsu.edu (S.E. Guttman).

0042-6989/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2006.09.012

structure refers to the overall pattern of timing with which events occur.

Theoretical considerations suggest that temporal structure, rather than synchrony, might represent the more reliable source of information for grouping. The probability of two unrelated events occurring synchronously, purely by chance, far outweighs the chance probability of obtaining two unrelated, identical *patterns* of events over time. Expressed in terms of information content (Shannon, 1948), common, irregular patterns over time convey more information than regular, synchronous events. Moreover, empirical studies of time-based grouping suggest that the nature of temporal structure affects task performance. Some experiments use *deterministic* temporal structure (e.g., Kandil & Fahle, 2001, 2004; Kiper, Gegenfurtner, & Movshon, 1996; Sekuler & Bennett, 2001): all stimulus elements change according to a regular, periodic pattern over time. In these studies, figure elements can be distinguished from ground elements solely on the basis of the phase of the changes (i.e., asynchrony). Other experiments present stimuli with *stochastic* temporal structure (e.g., Adelson & Farid, 1999; Guttman et al., 2005; Lee & Blake, 1999; Morgan & Castet, 2002): figure elements change at times designated by one stochastic process (i.e., elements are equally likely to change or to not change on any given frame), whereas background elements change at times designated by a different stochastic process. Under at least some conditions, stochastic temporal patterns yield more robust figure–ground segmentation than do deterministic temporal patterns (see review by Blake & Lee, 2005).

Temporal synchrony and temporal structure easily are confounded because multiple elements that undergo a series of synchronous changes have, by definition, the same temporal structure. The two can be distinguished, however, as we demonstrate here. In the current paper, we present psychophysical evidence that time-based grouping can be achieved based on patterns of temporal change. Specifically, we created animations in which temporal structure and temporal synchrony define opposing perceptual organizations and, using those animations, we compared the relative effectiveness of these two cues for grouping. The results from all experiments support the notion that temporal structure impacts perceived spatial organization more significantly than does temporal synchrony.

2. Experiment 1: Detecting asynchronies

The point of departure for this study was an anecdotal observation: when two Gabor patches changed asynchronously, detecting the temporal asynchrony seemed more difficult when those changes occurred multiple times (with constant temporal lag) during a viewing sequence. It was as if multiple samples of a given event paradoxically reduced the salience of the temporal asynchrony of those events.

Our first experiment was designed to evaluate this observation using more rigorous, forced-choice methodology. Is asynchrony more difficult to detect when changes occur within the context of a common temporal structure? If indeed verified, this finding could be leveraged into a test of the relative salience of temporal structure versus temporal synchrony in figure–ground segmentation.

2.1. Methods

2.1.1. Observers

Five observers with normal or corrected-to-normal vision participated in all experiments reported herein. Two observers were authors of this paper (SEG and LAG); the other three observers had previous experience with psychophysical observation but were naïve to the experimental hypotheses.

2.1.2. Apparatus

The stimuli for all experiments were generated with a Macintosh G4 computer and appeared on a gamma-corrected Mitsubishi Diamond Pro 2020u 20 inch monitor with a spatial resolution of 1280×1024 pixels and a refresh rate of 120 Hz. The monitor provided the only source of illumination in an otherwise darkened testing room.

2.1.3. Stimuli

The stimuli consisted of two, vertically arranged Gabor patches on a mid-gray (16.5 cd/m^2) background (Fig. 1A). Each Gabor patch had a visible area of approximately 0.80 deg (SD = 0.20 deg) and the center-to-center distance between elements measured 1.0 deg. Each Gabor patch had randomly assigned orientation, phase, spatial frequency (1.0–4.0 cycles/deg), and contrast (10–100%). During the course of a trial, each Gabor patch changed in spatial frequency once or multiple times; the new spatial frequency differed from the previous spatial frequency by at least 33% and fell within the range stated above, but otherwise was randomly determined. Previous work has shown that spatial frequency changes of this kind provide a highly salient cue for temporal grouping (Guttman et al., 2005), although those results do not speak to observers' sensitivity to asynchrony *per se*.

2.1.4. Procedure

The asynchrony detection experiments utilized a two-interval, forced choice procedure. In each of two intervals, observers viewed two Gabor patches that, at some point(s) in time, changed with respect to spatial frequency. In one interval, the changes occurred synchronously; in the other interval, one patch changed slightly earlier than did the other patch. The observer indicated, via keypress, the interval in which the spatial frequency changes occurred asynchronously.

In the first version of this task ("Single Change"), each element changed once within each interval of a trial (Fig. 1B). Specifically, the two Gabor patches appeared simultaneously for 500 ± 16.7 ms, followed by a single change in spatial frequency for each element. During one interval, the changes occurred synchronously; during the

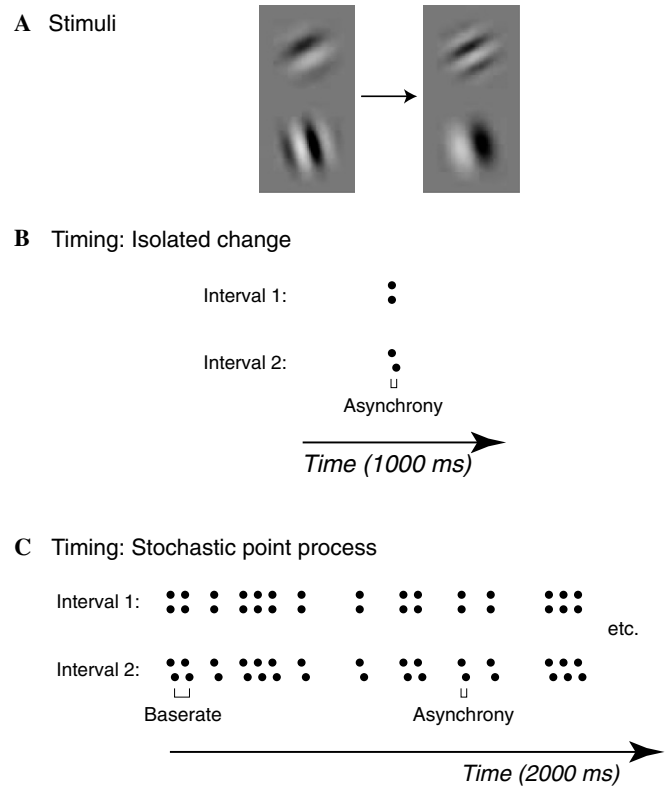

Fig. 1. Stimuli and trial timing for asynchrony detection experiments. (A) Two Gabor patches, presented during each interval of a 2-IFC procedure. Each element is random with respect to orientation, phase, contrast, and spatial frequency, and changes in spatial frequency at some point(s) in time. (B) Schematic depiction of trial timing for experiment in which observers must detect asynchrony when presented in the form of an isolated change. The dots depict the times at which the two Gabor patches changed. During each interval, the two elements appeared simultaneously, changed after approximately 500 ms, then disappeared simultaneously after a total exposure duration of 1000 ms. The relative asynchrony during the "asynchronous" interval (shown here in interval 2) varied across trials from 8.3 to 33.3 ms (1–4 frames at 120 Hz). (C) Schematic depiction of trial timing for experiment in which observers must detect asynchrony when presented in the context of a stochastic point process. Within a point process, the minimum time between two successive changes was 33.3 ms. The two elements changed a total of 30 times during each 2000 ms interval.

other interval, the changes occurred asynchronously, with the change of one patch lagging the other. The amount of temporal lag varied across trials from 8.3 to 33.3 ms (in 8.3 ms steps; 1–4 frames at 120 Hz), but was held constant within each trial. The entire array in each interval disappeared after 1000 ms.

In a second version of the task ("Stochastic Sequence"), the two Gabor patches each changed 30 times over 2000 ms according to a stochastic 30 Hz point process (i.e., every 33.3 ms, the patches would either change or not change with equal probability; Guttman et al., 2005). Within a trial, all patches changed according to the same point process; different stochastic sequences were used in different trials. In one interval the changes occurred synchronously; in the other interval, the two patches initially appeared simultaneously (and disappeared simultaneously at the end of the trial), but all changes of one Gabor patch lagged the changes of the other Gabor patch by a constant amount. As in the single change condition, the temporal lag varied across trials from 8.3 to 33.3 ms, in 8.3 ms steps.

In both versions of the task, the parameters of the Gabor patches both before and after change(s) were always the same in the two intervals, such that the intervals differed only in terms of synchronicity. The interval in which the changes occurred asynchronously varied randomly across trials, as did which patch (top or bottom) underwent the first change in that interval. Observers pressed one of two keys to indicate the interval in which the changes were asynchronous. For each task, observers participated in 256 randomly ordered trials over four sessions, for a total of 64 trials at each level of asynchrony.

2.2. Results and discussion

Fig. 2 depicts the proportion of responses correct as a function of the amount of asynchrony. The results of the single-change experiment indicate that human observers can effectively distinguish synchrony from asynchrony. When the stimulus changes were conveyed as isolated events, task performance increased significantly with the amount of asynchrony, $F(1,4) = 367.0$, $p < .001$. Still, discrimination performance, when averaged across observers, actually exceeded chance levels for all measured asynchronies, $t_4 \geqslant 5.3$, $p < .01$. Furthermore, 4 of 5 individual observers significantly exceeded chance discrimination levels for single-change asynchronies at and above 16.7 ms, $\chi^2 \geqslant 4.0$, $p < .05$. Thus, the results of the single-change experiment reveal that human observers can detect asynchronous stimulus change on the order of 8–16 ms. This estimate compares favorably with previously measured thresholds for temporal order detection (e.g., Hirsh et al., 1961; Westheimer & McKee, 1977).

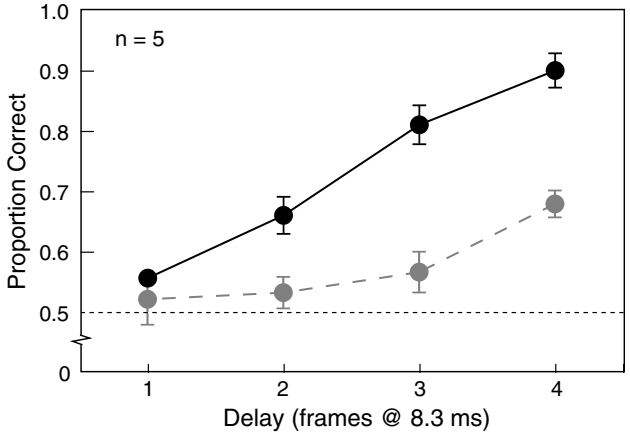

Fig. 2. Asynchrony detection as a function of amount of asynchrony. Black symbols indicate the proportion of correct responses when asynchronies were presented as isolated events. Gray symbols indicate the proportion of correct responses when asynchronies were embedded in stochastic point processes. Error bars represent ±1 SE across observers.

Consistent with our anecdotal observations, asynchrony detection was indeed more difficult when the stimulus changes were embedded in stochastic sequences (see Fig. 2). As before, task performance improved with the amount of asynchrony, $F(1,4) = 111.2$, $p < .001$. However, under these conditions observers required 33.3 ms of asynchrony before task performance significantly exceeded chance levels, $t_4 = 5.0$, $p < .01$. Overall, observers distinguished the synchronous from the asynchronous interval significantly less effectively when presented with 30 changes in the context of stochastic point processes compared to the single presentation of one change, $F(1,4) = 38.3$, $p < .01$, particularly at higher levels of asynchrony (for interaction, $F(3,12) = 4.4$, $p < .05$).

At first glance, this finding seems counterintuitive. We know that repetitive temporal asynchrony has perceptual consequences: out-of-phase flicker among spatially segregated clusters of elements can promote figure–ground segmentation (Sekuler & Bennett, 2001), even when the flicker itself is indistinct (Rogers-Ramachandran & Ramachandran, 1998). So why should it be more difficult to perceive asynchrony when viewing stochastic point processes rather than single changes? Why, in other words, does a greater number of relevant events produce poorer discrimination? Perhaps when observers view an extended sequence of stochastic events, the temporal structure (i.e., pattern) of those events is more salient than the temporal synchrony (or lack thereof) among events within the sequence. If this were true, the extended sequence could create a cue conflict situation in which the similarity in temporal pattern (structure) dominates otherwise detectable differences in the absolute timing (synchrony) of individual changes. This context dependence of asynchrony detection suggested to us that temporal structure may play an important role in visual grouping, a role not revealed in earlier work using repetitive flicker. This possibility was tested in the following experiments.

3. Experiment 2: Opposing perceptual organizations

In this series of experiments, we tested directly the nature of the time-based cue underlying spatial grouping and segmentation. Specifically, we asked whether temporal structure or temporal synchrony would dominate perceived grouping when the two defined opposing perceptual organizations.

3.1. Method

Observers viewed 6×6 arrays of Gabor patches (Fig. 3A). As in Experiment 1, each Gabor patch had a visible area of approximately 0.80 deg and the elements were separated by 1.0 deg (center-to-center distance), creating an overall stimulus measuring just under 6.0 deg. All Gabor patches had randomly assigned orientation, phase, spatial frequency, and contrast, within the same ranges as used in Experiment 1.

Over the course of two seconds, each element within the array changed in spatial frequency 30 times according to one of two stochastic point processes (Fig. 3B). Specifically, the central four elements changed in spatial frequency at times designated by the "figure" point process. Two sets of flanking elements also changed according to the figure point process, but with all changes delayed by 1–4 frames (8.3–33.3 ms) relative to the central elements; whether these flankers fell to the left and right or above and below the central elements varied across trials. The other two sets of flanking elements changed at times designated by the second, "ground" point process. The correlation between the figure and ground point processes varied from 0 to 0.8, with higher correlations providing less information to segregate figure from ground (i.e., the two regions were more similar in temporal pattern); previous research confirms that this manipulation significantly influences the strength of figure–ground segregation (Guttman et al., 2005). Finally, the remaining elements changed at times designated by a version of the ground point process that was delayed by the same amount as the figure delay for that trial. The task was to determine whether the elements that more strongly grouped together—and segregated from the rest of the array—formed a horizontal or vertical rectangle. Observers pressed one of two keys to indicate their

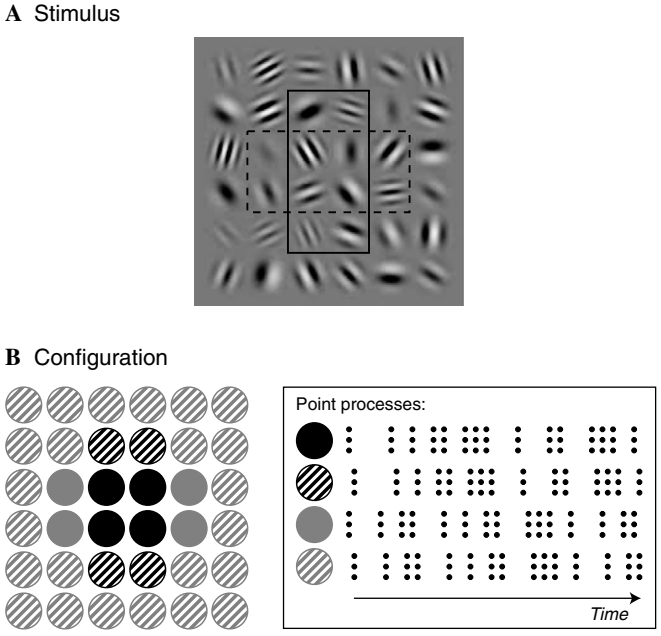

Fig. 3. Stimuli and trial timing for perceptual organization experiment. (A) Array of Gabor patches seen during a single frame of a trial. The solid and dotted rectangles depict the two possible perceptual organizations between which observers must choose. (B) Schematic illustration of stimulus configuration, in which different elements change according to different point processes. The dots at the right of the figure depict the times at which elements in each of the four categories changed; all elements underwent 30 changes over the course of 2000 ms. Solid black circles depict elements that change according to the "figure" point process. Striped black circles depict elements that change according the figure point process, but delayed relative to the central figure elements. Solid gray circles depict elements that change according to the "ground" point process. Striped gray circles depict elements that change according a delayed version of the ground point process.

judgment; instructions indicated that one of the two orientations must be chosen, even if no clear figure popped out or if the orientation appeared to fluctuate over the course of the trial.

The total trial length varied between 2041.7 and 2066.7 ms, consisting of an initial 33.3 ms static frame, 2000 ms of changes, and a temporal lag in the delayed point processes of 8.3–33.3 ms. All elements appeared and disappeared simultaneously at the beginning and end, respectively, of each trial. For the initial experiment, observers participated in 640 randomly ordered trials over eight sessions, resulting in 32 observations at each level of asynchrony and figure–ground correlation. In a follow-up experiment (described below), observers participated in 320 trials over four sessions, for a total of 64 trials at each level of correlation.

3.2. Results and discussion

If spatial grouping and segmentation are strongly dependent on temporal structure, then observers should group elements within an array that change according to the same point process, even when there exist timing delays among those elements; this grouping should cause observers to segment the array based on the different point processes (i.e., "vertical" response for Fig. 3B). On the other hand, if grouping is more strongly dependent on temporal synchrony, then observers should group elements that often change at the same absolute times (as the "figure" and "ground" elements do, particularly at higher correlations), ignoring differences in the global temporal structures; perceptual organization based on temporal synchrony would, therefore, favor segmentation based on the differences in absolute timing (i.e., "horizontal" response for Fig. 3B).

Plotted in Fig. 4 is the proportion of trials on which observers chose the figure orientation consistent with grouping by temporal structure for the various combinations of figure–ground correlation and delay. Observers responded "structure" less frequently as correlation increased, $F(4,16) = 194.5$, $p < .001$; this result is not surprising, as higher correlations correspond to less structural difference to distinguish figure from ground. The proportion of "structure" responses also decreased as a function of increasing delay, $F(3,12) = 62.1$, $p < .001$; consistent with previous findings (e.g., Fahle & Koch, 1995), asynchrony appears to play a larger role in determining perceived spatial organization as the amount of asynchrony increases. Most importantly, however, the proportion of trials on which observers responded "structure" was above 0.5 for the overwhelming majority of conditions. Not until figure–ground correlation rose to 0.8 (i.e., very little temporal structure difference was available to distinguish "figure" from "ground") and the asynchrony within a region was four frames (33.3 ms) did observers group by temporal synchrony more often than by temporal structure. In sum, observers systematically grouped elements that changed according to the same pattern over time—even though the changes occurred asynchronously across elements—and segregated elements that changed according to different patterns.

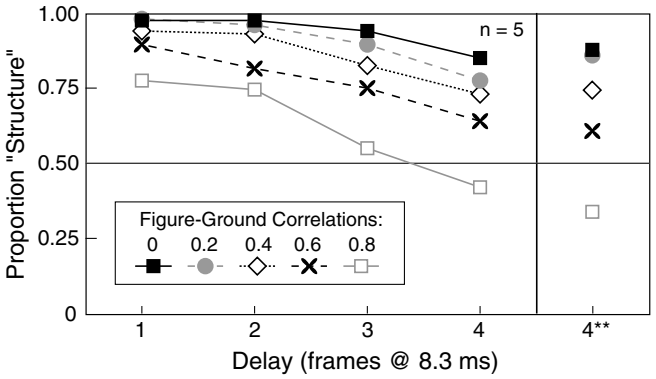

Fig. 4. Perceived organization when temporal structure opposes temporal synchrony: Proportion of responses consistent with grouping by temporal structure as a function of amount of asynchrony (delay) and figure–ground correlation. The left panel depicts the experiment in which delay varied across trials; the right panel (indicated by 4**) depicts the follow-up experiment in which delay was set to four frames (33.3 ms), and both the figure and ground point processes were controlled to have zero correlation with their delayed versions. Responses above 0.5 indicate grouping and segmentation on the basis of temporal structure, whereas responses below 0.5 indicate grouping and segmentation on the basis of temporal synchrony. To minimize visual clutter, error bars have been omitted from this graph; across conditions, standard error ranged from 0.008 to 0.054, averaging 0.026.

In a follow-up experiment, we replicated the 4-frame delay condition with additional control over temporal synchrony. The delay of four frames (33.3 ms) is highly detectable (Experiment 1) and matches the time between potential changes within a single point process, such that a change within the delayed point process may co-occur with a non-delayed change from the fundamental point process. Here, the figure and ground point processes were selected such that each point process, relative to its delayed version, had zero correlation. In this manner, the level of synchrony *within* a "grouping-by-structure region" was always quantitatively the same or less than the level of synchrony between regions. Thus, if synchrony dominates perceptual grouping, then observers should ignore the different temporal structures and consistently select the organization in which the asynchronous point processes define different regions.

In contrast to this prediction, the results of this experiment did not differ qualitatively from the corresponding conditions of the experiment described above (right panel of Fig. 4). Despite the additional control over temporal synchrony, observers still showed a strong tendency to organize the array in accordance with temporal structure.

In sum, temporal structure, rather than temporal synchrony, dominates perceived grouping and segmentation, at least for the range of delay values tested in this experiment. In determining the perceptual organization of a dynamic array, human vision appears to rely more on

patterns of change over time than on the absolute timing of events that comprise the pattern.

4. Experiment 3: Objective grouping task

These conclusions regarding the relative importance of temporal structure and temporal synchrony, although based on highly systematic data, depend on subjective judgments about the perceived shape of a figure seen against a background. To eliminate any subjectivity, we designed an objective task to probe the relative efficacy of temporal structure and temporal synchrony for supporting spatial grouping and segmentation of dynamic arrays; performance on this modified task was tested in our third experiment.

4.1. Method

To ensure adequate task difficulty (i.e., off-ceiling performance), several stimulus changes were required. Observers viewed a 20×20 array of stochastically changing Gabor patches in which each patch changed according to one of two 30 Hz stochastic point processes (Fig. 5A). For this experiment, the Gabor patches had a visible area of approximately 0.67 deg (SD = 0.17 deg) and the distance between elements measured 0.80 deg, for a total stimulus size of approximately 16.0 deg. As before, each element had a different, random orientation, phase, spatial frequency, and contrast, but spatial frequency ranged from 0.67 to 2.67 cycles/deg and contrast ranged from 5 to 30%. (Note that these ranges of spatial frequency and contrast values were considerably smaller than those used in Experiments 1 and 2, to make the task more difficult.)

The "figure" consisted of a single 10×12 rectangle, oriented either horizontally or vertically, and positioned randomly in the array with the constraint that at least two patches had to intervene between the figure and all edges of the array. Within the figure region, all elements changed in spatial frequency at times designated by one point process. The remaining "ground" elements changed according to the second point process. The point processes operated at the same rate as in previous experiments (30 Hz; changes occurring at random multiples of 33.3 ms), but for a total of 10 changes over a 700 ms trial (initial 33.3 ms frame +666.7 ms of changes).

To determine the importance of temporal structure, the relationship between the figure and ground point processes varied across trials (Fig. 5B). The figure and ground point processes could be: (1) independent; (2) the same, but delayed by four frames (33.3 ms) relative to one another; or (3) the same, but delayed by eight frames (66.7 ms) relative to one another. All resulting figure–ground combinations had a temporal correlation of zero, such that the number of asynchronous changes distinguishing the figure from the ground was the same across conditions. Thus, any differences in performance across conditions must be attributed to variations in the temporal structure relationship between the figure and ground regions. Observers participated in 300 trials over four sessions, resulting in 100 trials for each type of point process.

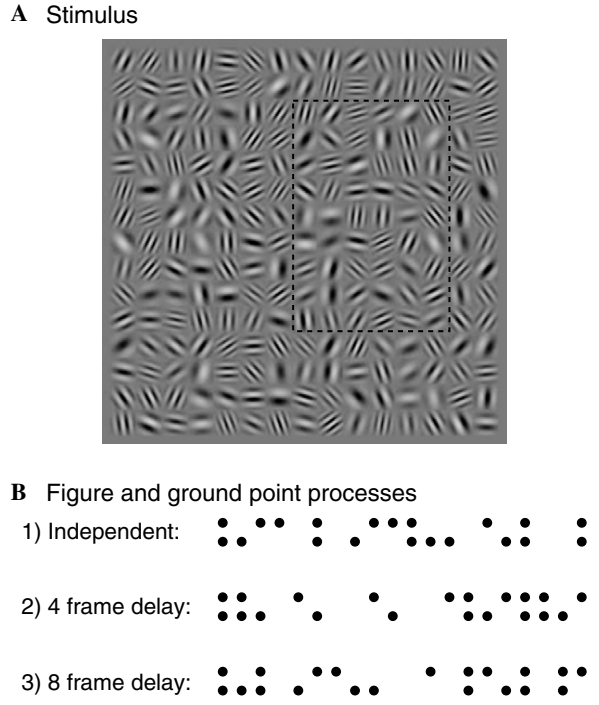

Fig. 5. Stimuli and trial timing for objective grouping experiment. (A) Array of Gabor patches seen during a single frame of a trial. The dotted rectangle indicates a possible figure region, here depicted vertically. In the actual experiment, the stimulus contained 20×20 elements and the figure region contained 10×12 elements. (B) Examples of point processes defining the timing with which the stimulus elements changed. The top row in each pair represents the "figure" point process and the bottom row represents the "ground" point process. (1) Independent: the two point processes are independently stochastic. (2) Four frame delay: the two point processes have the same temporal structure, but with the ground elements changing four frames (33.3 ms) later than the figure elements. (3) Eight frame delay: the two point processes have the same temporal structure, but with the ground elements changing eight frames (66.7 ms) later than the figure elements. For the delayed point processes, the final one or two elements of the original point process were wrapped around to the beginning of the sequence, if necessary.

4.2. Results and discussion

Regardless of the relationship between the figure and ground point processes, the times at which figure elements changed were completely uncorrelated with the times at which ground elements changed. Therefore, if perceptual organization depends on temporal synchrony, then observers should exhibit similar task performance in all three conditions. If, however, perceptual organization depends on temporal structure, then task performance should be better in the independent condition than in the 4-frame delay condition. This prediction follows from the fact that distinct temporal structures distinguish figure and ground in the independent condition, whereas the figure and ground regions in the 4-frame delay condition have the same tem-

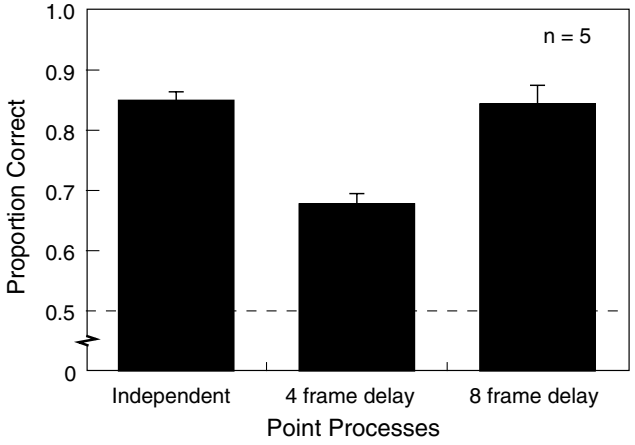

Fig. 6. Figure detection as a function of figure–ground relationship. Proportion correct horizontal–vertical judgments as a function of the relationship between the figure and ground point processes. Error bars represent 1 SE across observers.

poral structure and thus can be distinguished only on the basis of asynchrony. We further would predict that the 8-frame delay condition should produce comparable performance to the independent condition, as the delay of eight frames essentially produces qualitatively different temporal structures to distinguish figure and ground regions.

Plotted in Fig. 6 is the proportion of correct responses as a function of the relationship between the figure and ground point processes. Although task performance exceeded chance levels for all conditions ($t_4 \geqslant 10.6$, $p < .001$), the effectiveness of grouping and segmentation clearly varied with the figure–ground relationship. Specifically, observers segregated figure from ground less effectively when the two regions had the same temporal structure (i.e., 4-frame delay), even though the level of synchrony vs. asynchrony was the same. The results of this experiment support the conclusions that (1) temporal structure provides the more potent cue for segmentation, and (2) the time-based information for grouping cannot be reduced to temporal synchrony.[1] Results from the 8-frame delay imply that there is an upper limit to the temporal precision with which identical but delayed point processes can be correlated, as one would expect. Further work is required to establish the exact temporal constraints on the efficacy of temporal structure.

5. Experiment 4: Does "jitter" disrupt grouping based on pattern?

One could argue that observers are not sensitive to the pattern of changes over time (temporal structure) but, instead, are picking up on relatively coarse temporal information engendered by elements defining the figure region. Indeed, an appropriately designed spatiotemporal filter could register temporal correlation over any arbitrarily long time scale. To test this hypothesis, we compared segmentation and grouping performance using jittered point processes to segmentation and grouping performance using delayed point processes, with the two arrays having the same number of correlated events within a given time window. If observers are relying on coarse temporal correlations and not temporal structure, the jittered point processes and delayed point processes should have comparable effects on perceptual grouping.

5.1. Method

Experiment 4 used dynamic arrays of Gabor patches identical to those of Experiment 3, except that contrast ranged from 10 to 100% for added visibility in this difficult task. Fig. 7A illustrates the point processes defining figure and ground in these two categories of stochastic displays. To create a given animation, we started with a 30 Hz stochastic point process like those used in our other experiments; as in Experiment 3, each element underwent 10 changes over the course of 666.7 ms. This point process defined the times at which the figure elements, a 10×12 region within a 20×20 array of Gabor patches, changed

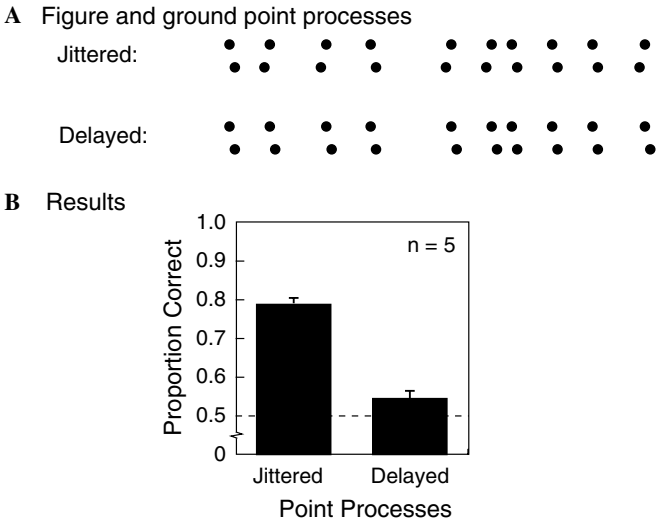

Fig. 7. Jittered versus delayed point processes. (A) Examples of point processes defining the timing of changes in figure and ground elements. In the "delayed" condition, all change points associated with figure elements occur one frame after the change points for ground elements (or vice versa). In the "jittered" condition, change points associated with figure elements occur either one frame before or one frame after the change points for ground elements. Every jittered sequence contained five instances where figure elements changed before ground elements and five instances where the figure elements changes after ground elements, with these two types of change randomly intermixed over the entire sequence of 10 changes. (B) For each of five observers, percent correct on the two-alternative, forced-choice task (chance equals 50%) for the two types of animations, jittered and delayed. Error bars represent 1 SE across observers.

[1] One cannot directly compare results for the independent condition in Experiment 3 (~85% correct, Fig. 6) with the zero-correlation conditions from Experiment 2 (~94% correct, Fig. 4), because the displays and task are quite different in the two experiments.

in spatial frequency. For the jittered condition, the background elements in the array were assigned the same point process as the figure elements, except that the points in time at which a change occurred was shifted either one frame (8.3 ms) in advance of the figure change times (''lead'' change) or one frame after the figure change times (''lag'' change). There were ten change points in total within a given animation; we ensured that five of those were ''lead'' changes and the other five were ''lag'' changes, with the order of leads and lags being random. For the delayed condition, all stimulus elements within the background region changed either one frame in advance of the background elements or one frame after the background elements; in other words, one set of changes was always delayed relative to the other. Note that the amount of asynchrony between figure and ground was the same for both jittered and delayed displays. Note also that both jitters and delays introduce an 8.3 ms asynchrony between each figure change and the ground change that occurred most closely in time, the smallest lag possible within the limits of a 120 Hz video monitor.

Observers performed the same task as in our other experiments: judging whether the figure region had a horizontal or a vertical orientation. Each of our five observers were tested on 200 trials, administered in blocks of 50, with jittered and delayed animations randomly intermixed.

5.2. Results and discussion

Consider the stimulus information potentially available for performing this task. Jittered and delayed displays contain equivalent amounts of temporal offset information to distinguish figure from ground, which implies that performance based on these two categories of displays should be comparable if observers use asynchrony information to segregate a figure. But if observers rely on temporal pattern, then the two types of displays should produce different levels of performance. Specifically, with the delayed point processes, both figure and ground elements change according to the same pattern, which should make stimulus segmentation difficult if observers rely on this source of information to segregate figure from background. With the jittered point process, however, figure and ground elements change according to different patterns and, therefore, should be discriminable.

The results of this experiment support the latter prediction. All five observers tested on this task identified figure orientation with reliably greater accuracy in displays with jittered point processes compared to displays with delayed point processes, $t_4 = 8.6$, $p < .001$ (Fig. 7B). Indeed, relying on delay to segregate figure from ground resulted in essentially chance performance, which is not so surprising in light of the results from Experiment 1 showing that the difference between asynchronous point processes is very difficult to judge. We take this finding as further evidence that common temporal structure promotes grouping which, in this experiment, impaired the ability to segregate figure from ground because elements in both regions had the same temporal structure.

6. Temporal structure and spatiotemporal filtering

Previous conclusions from work on time-based grouping (Lee & Blake, 1999) have been criticized for ignoring the possible role of spatiotemporal filtering (such as that implemented by motion energy models) in the extraction of form from temporal structure. Thus, Adelson and Farid (1999) and, subsequently, Farid (2002) showed that physiologically plausible neural filters with appropriately selected temporal bandpass characteristics could reveal spatial structure in the stochastic displays developed by Lee and Blake and used, with modification, in the present experiments. This filter response occurs because, in stochastic displays, there are periods during which stimulus elements defining the figure create abrupt transient signals while stimulus elements defining the background do not. These transient events can be readily detected by neural filters with biphasic responses, or transient detectors as Lee and Blake called them. As Farid and Adelson (2001) correctly argue, these coarse-scale temporal changes can be registered by temporal bandpass filters without recourse to the fine temporal resolution required to encode temporal synchrony (see also Morgan & Castet, 2002).

Elsewhere, we have acknowledged the possible role of transient detectors in registering form from temporal structure (Guttman et al., 2005), and we *a priori* have no reason to doubt the involvement of transient detectors in the grouping effects described here. Indeed, it is natural to ask what impact the current class of temporal structure displays would have on these putative biphasic filters. Is there spatial structure arising within the filtered outputs of the animations used in this study, and, if so, do those outputs predict the pattern of empirical results when temporal synchrony opposes temporal structure?

To answer those questions, we implemented the temporal bandpass filter described by Farid and Adelson (2001) and Farid (2002), in which the bandpass impulse response is given by:

$$h(t) = (kt/\tau)^n \mathrm{e}^{-kt/\tau}\left[1/n! - (kt/\tau)^2/(n+2)!\right]$$

with $\tau = 0{:}01$, $k = 2$, $n = 4$, and an integration time covering 100 ms (coinciding with the time constant used by Farid and Adelson and by Farid). We applied this filter to a sample of stimulus displays used in Experiment 2, in which temporal structure defined one spatial grouping (e.g., ''horizontal rectangle'') and temporal synchrony defined a different spatial grouping (e.g., ''vertical rectangle'').

For each animation, the 84-frame original was filtered through the biphasic filter, producing a series of matrices each containing values proportional to the output of the filter at a given spatial location. Examples of a successive series of these filtered outputs are shown in Fig. 8A, in exactly the same format utilized by Farid and Adelson (2001) and

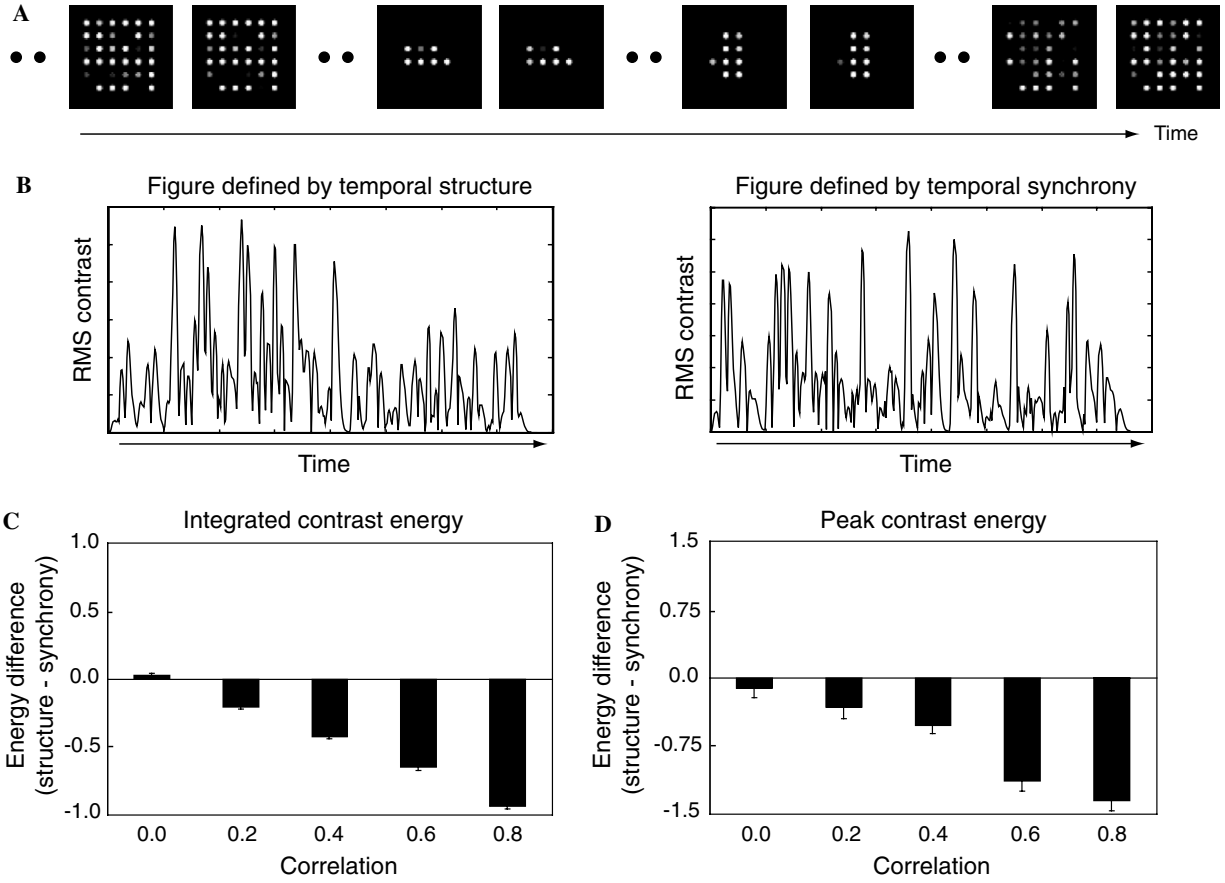

Fig. 8. Dissociation of perceived organization and spatiotemporal filtering models. (A) Sample "frames" of the output sequence resulting from the convolution of a temporal bandpass (biphasic) filter with an animation sequence used in the present study. Often, no clear spatial structure is evident in the filtered output (e.g., the first and last pair of frames), but occasionally there are brief moments in which horizontal or vertical structure is clearly evident (i.e., the middle pairs of frames). (B) Fluctuations in RMS contrast corresponding to the spatial form defined by temporal structure (left panel) and by temporal synchrony (right panel). (C) The difference in integrated contrast energy for structure- and synchrony-defined form as a function of the correlation between figure and ground point processes (there were 80 exemplars for each correlation value; error bars denote ±1 SE). Positive values indicate that contrast energy favors structure-defined form and negative values indicate that contrast energy favors synchrony-defined form. (D) The difference in peak contrast energy for structure- and synchrony-defined form (positive and negative values correspond to structure- and synchrony-defined form, respectively).

by Farid (2002). Looking at all the filtered sequences, one often sees no spatial structure in the filtered output, but one can find brief moments in time in which the filtered output clearly signals temporally correlated events within a region of the array. Significantly, that structured region sometimes corresponds to the region defined by temporal synchrony and other times it corresponds to the region defined by temporal structure. To quantify any bias toward structure- versus synchrony-defined form, we computed the RMS contrast between "figure" and "ground" pixels contained in each frame of the filtered animations. This index was computed separately for the pixels specifying the form associated with temporal structure and for the pixels specifying the form associated with temporal synchrony. We thus generated two alternative records of fluctuations in RMS contrast for each and every filtered sequence. An example of those fluctuations for these two alternative definitions of form are shown in Fig. 8B. Note the considerable variability in the strength of the contrast between figure and ground pixels in both plots. On some frames, the "synchrony" defined contrast dominates and on other frames the "structure" defined contrast dominates.

It is important to keep in mind that these images of the filtered outputs of our animations portray on a pixel by pixel basis the output levels, over space and time, of an array of biphasic filters. These images do not specify how neighboring responses become grouped over space to recover a given shape. Furthermore, it is not possible to specify the minimum contrast necessary to support perception of the spatial structure implicitly contained in these sequences of outputs. Thus to compare the contrast signals associated with temporal structure and with temporal synchrony, we have made the very unrealistic assumption that *any* non-zero contrast value could specify a given form.

To predict the perceptual judgment produced by a given filtered sequence, we implemented two alternative decision rules, one based on integrated contrast energy and the other based on peak contrast energy. For the integration rule, we integrated over the entire animation sequence the

contrast energy associated with the form defined by structure and subtracted from that the integrated contrast energy associated with the form defined by synchrony. Fig. 8C shows the distributions of those difference values for the five different levels of correlation employed in Experiment 2; Fig. 8D shows the comparable distributions when contrast is quantified in terms of the peak contrast value in each sequence. For both decision rules, contrast energy tends to favor form defined by synchrony for all correlation values except zero correlation.

Comparing panels C and D of Fig. 8 with the actual data from Experiment 2 reveals that the biphasic filter model does not provide reliable predictions of performance. On trials in which the correlation between the timing of figure and ground changes is zero, observers were strongly biased to perceptually organize the display by temporal structure, not temporal synchrony. However, the outputs of the hypothetical filters show essentially no bias; indeed, for the zero correlation condition, the model predicts an approximately equal distribution of category responses for the two alternatives specified by structure and by synchrony. And at higher values of correlation, the model predicts bias in favor of synchrony when, in fact, observers continue to report form defined by structure. Only when correlation approaches unity does perceptual organization become defined by synchrony, as predicted by the biphasic filter model.

One could argue, of course, that our selection of the time constant for the filter unwittingly biased results against the structure solution. We specifically used the parameter value utilized in most of Farid and Adelson's work, where the filter was able to recover spatial structure in stochastic displays that confounded temporal synchrony and temporal structure. Still, to test the generality of our analyses, we created new filtered outputs using a much longer time constant, 300 ms. Results from that reanalysis produced essentially the same pattern of predictions, with the filtered outputs either showing no bias for synchrony versus structure, or a bias for synchrony when, in fact, structure dominated in perception.

This analysis does *not* imply that biphasic filters are irrelevant for registering temporal structure. On the contrary, we can imagine no other way that the visual system could register the rapid, irregular changes in spatial frequency that define spatial structure in these stochastic displays. Moreover, we agree with Farid (2002) that dynamic displays producing negligible coherent responses in an array of biphasic filters are likely to produce weak or nonexistent spatial grouping. What the analyses in this section imply to us is that spatial structure in stochastic displays is not being recovered by coarse temporal correlation but, instead, by a temporal pattern matching process that does not require precise temporal synchrony for registration of common temporal structure. The details of this pattern matching process remain to be learned, but the results from the current study suggest that temporal structure provides the key to grouping in these displays.

7. General discussion

7.1. Temporal structure and its relation to previous psychophysical studies

Previous psychophysical work supports the notion that time-based cues, in general, lead to spatial grouping and segmentation. Several of these studies found effective grouping using deterministic temporal structure (i.e., figure and ground regions changed according to the same, periodic schedule), meaning that different regions were distinguishable solely on the basis of temporal synchrony (e.g., Kandil & Fahle, 2001; Rogers-Ramachandran & Ramachandran, 1998; Sekuler & Bennett, 2001; Usher & Donnelly, 1998).

The ideas presented herein are entirely consistent with these previous psychophysical studies. To be sure, our results do not imply that synchrony fails to influence perceptual organization; when choosing between alternative perceptual organizations, observers in our Experiment 2 showed an increased tendency to group elements on the basis synchrony when temporal structure was purposefully degraded. Nor do our results rule out an interaction between temporal synchrony and temporal structure in determining perceived grouping. However, when temporal structure was reliably present in the stimulus array, that cue—not synchrony—dominated.

Interestingly, two aspects of the earlier research support the notion that temporal structure provides a stronger cue for grouping and segmentation than does temporal synchrony. First, segmentation seen in deterministic displays may be attributed largely to asynchronies in initial stimulus onset (Beaudot, 2002); repeated presentations of asynchronous events (i.e., out-of-phase flicker) actually weakens perceptual organization (Kandil & Fahle, 2001; Usher & Donnelly, 1998). In other words, the equivalence of the temporal patterns in the figure and ground regions tended to overwhelm differences in absolute timing, just as we found in the current study. Second, studies using deterministic temporal structure—wherein asynchrony is the only temporal cue distinguishing figure from ground—have found that synchrony fails to influence perceptual organization when the displays carry meaningful spatial information (Fahle & Koch, 1995; Kiper et al., 1996; cf. Usher & Donnelly, 1998). By contrast, stochastic temporal structure interacts synergistically with spatial structure in determining perceptual organization (Lee & Blake, 2001). In this regard, it is noteworthy that our study eliminated spatial cues for grouping altogether, and the results clearly supported temporal structure, rather than temporal synchrony, as the temporal cue of greatest salience.

The current study also dovetails nicely with results reported by Guttman et al. (2005), whereby observers grouped elements that changed at times designated by the same stochastic process even when those elements changed in very different ways (i.e., elements changing in contrast readily grouped with elements changing in spatial frequency). At

first glance this result seems counterintuitive because the latencies with which the visual system signals change vary significantly with spatial frequency and contrast, not to mention the nature of the change. Grouping based on temporal synchrony would be confounded by the absolute latency differences in registering these kinds of diverse stimulus elements. Grouping based on temporal patterns of change over time, however, would be robust across modest differences in absolute latency.

In sum, several lines of evidence, including those presented herein, suggest that temporal structure provides a more salient cue for perceptual organization than does temporal synchrony. Future work is needed to determine the level of temporal precision required to establish common temporal structure, as well as the length of the sequence needed to achieve binding.

7.2. Temporal structure: A key to neural binding?

Time-based visual grouping has been interpreted by some as evidence for the *temporal correlation hypothesis*—the idea that the visual system implements feature binding via temporally correlated responses among neurons that encode elements belonging to a single object (Eckhorn, 2000; Engel & Singer, 2001; Singer & Gray, 1995). But in what manner must these neural impulses be correlated? The psychophysical results presented here indicate that temporal structure—patterns of stimulus change over time—supports stimulus grouping and segmentation more effectively than does temporal synchrony *per se*, at least within the tested range of temporal delays. By extension, we predict that patterns of neural firing over time, irrespective of the precise millisecond timing of individual spikes, may constitute an important signature of neural binding. This *temporal structure hypothesis* is consistent with the suggestion that correlated neural activity underlies binding. At the same time, the hypothesis sidesteps concerns that the visual system cannot maintain precise synchrony through multiple layers of processing (Mainen & Sejnowski, 1995; Roelfsema, Lamme, & Spekreijse, 2004; Shadlen & Movshon, 1999). According to the scheme proposed here, precise synchrony is unnecessary because grouping relies on correlated "rhythms" of spike activity, within which some degree of timing noise can be tolerated. This hypothesis also does not require that stimulus events be registered at unrealistically fine temporal resolution, a reasonable criticism that has been leveled against the temporal synchrony hypothesis (Morgan & Castet, 2002). According to the temporal structure hypothesis, binding is contingent primarily on the reliability of patterns of neural activity produced by given sequences of stimulus events, and there is solid evidence for such reliability in neural spike trains (Bair & Koch, 1996; Berry, Warland, & Meister, 1997; Mainen & Sejnowski, 1995). In general, we are led to speculate that it may be biologically more plausible for the visual system to promote grouping based on temporal patterns of stimulus events (and, hence, temporal patterns of neural spikes), rather than on the absolute timing of stimulus events (and, hence, the absolute timing of individual neural spikes).

It remains to be learned how common temporal structure across an array of visually activated neurons could be registered, but algorithms for accomplishing this kind of operation have been described in other domains (Mozer, 1995). We hope that the psychophysical results presented here will motivate neurophysiological work to determine whether neural systems analogously depend on patterns of firing over time to encode the binding of local visual features into unified, global objects.

Acknowledgments

We thank David Bloom, Daniel Kaali, and Chai-Youn Kim for participating in these experiments. We are grateful to anonymous reviewers of an earlier version of this paper for their useful comments. This research was funded by NIH Grant EY07760.

References

Adelson, E. H., & Farid, H. (1999). Filtering reveals form in temporally structured displays. *Science, 286*, 2231a.

Alais, D., Blake, R., & Lee, S.-H. (1998). Visual features that vary together over time group together over space. *Nature Neuroscience, 1*, 160–164.

Bair, W., & Koch, C. (1996). Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Computation, 8*, 1185–1202.

Beaudot, W. H. A. (2002). Role of onset asynchrony in contour integration. *Vision Research, 42*, 1–9.

Berry, M. J., Warland, D. K., & Meister, M. (1997). The structure and precision of retinal spike trains. *Proceedings of the National Academy of Sciences USA, 94*, 5411–5416.

Blake, R., & Lee, S.-H. (2005). The role of temporal structure in human vision. *Behavioral and Cognitive Neuroscience Reviews, 4*, 21–42.

Eckhorn, R. (2000). Cortical synchronization suggests neural principles of visual feature grouping. *Acta Neurobiologiae Experimentalis, 60*, 261–269.

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences, 5*, 15–25.

Fahle, M., & Koch, C. (1995). Spatial displacement, but not temporal asynchrony, destroys figural binding. *Vision Research, 4*, 491–494.

Farid, H. (2002). Temporal synchrony in perceptual grouping: a critique. *Trends in Cognitive Sciences, 6*, 284–288.

Farid, H., & Adelson, E. H. (2001). Synchrony does not promote grouping in temporally structured displays. *Nature Neuroscience, 4*, 875–876.

Guttman, S. E., Gilroy, L. A., & Blake, R. (2005). Mixed messengers, unified message: spatial grouping from temporal structure. *Vision Research, 45*, 1021–1030.

Hirsh, I. J., & Sherrick, C. E. Jr., (1961). Perceived order in different sense modalities. *Journal of Experimental Psychology, 62*, 423–432.

Kandil, F. I., & Fahle, M. (2001). Purely temporal figure–ground segregation. *European Journal of Neuroscience, 13*, 2004–2008.

Kandil, F. I., & Fahle, M. (2004). Figure-ground segregation can rely on differences in motion direction. *Vision Research, 44*, 3177–3182.

Kiper, D. C., Gegenfurtner, K. R., & Movshon, J. A. (1996). Cortical oscillatory responses do not affect visual segmentation. *Vision Research, 36*, 539–544.

Lee, S.-H., & Blake, R. (1999). Visual form created solely from temporal structure. *Science, 284*, 1165–1168.

Lee, S.-H., & Blake, R. (2001). Neural synergy in visual grouping: when good continuation meets common fate. *Vision Research, 41*, 2057–2064.

Mainen, Z. F., & Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science, 268*, 1503–1506.

Morgan, M., & Castet, E. (2002). High temporal frequency synchrony is insufficient for perceptual grouping. *Proceedings of the Royal Society of London B, 269*, 513–516.

Mozer, M. C. (1995). Neural net architectures for temporal sequence processing. In A. S. Weigend & N. A. Gershenfeld (Eds.), *Time series prediction: Forecasting the future and understanding the past*. Redwood City, CA: Addison Wesley.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nature Neuroscience, 7*, 982–991.

Rogers-Ramachandran, D. C., & Ramachandran, V. S. (1998). Psychophysical evidence for boundary and surface systems in human vision. *Vision Research, 38*, 71–77.

Sekuler, A. B., & Bennett, P. J. (2001). Generalized common fate: grouping by common luminance changes. *Psychological Science, 12*, 437–444.

Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron, 24*, 67–77.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379–423, 623–656.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience, 18*, 555–586.

Suzuki, S., & Grabowecky, M. (2002). Overlapping features can be parsed on the basis of rapid temporal cues that produce stable emergent percepts. *Vision Research, 42*, 2669–2692.

Usher, M., & Donnelly, N. (1998). Visual synchrony affects binding and segmentation in perception. *Nature, 394*, 179–182.

Westheimer, G., & McKee, S. P. (1977). Perception of temporal order in adjacent visual stimuli. *Vision Research, 17*, 887–892.